

News from the Technology Transfer Program

Searching the future enterprise

NSA researcher invents system for efficient "collision-free hashing of near-match inputs"

s data services continue to expand, the problems involved in Big Data searches will increase. An invention by an NSA researcher offers a way to reduce the problem of "data collision" in enterprise-level systems.

People who manage massive databases know it's important to ensure that data relationships stay clear. When pieces of data exist in many-to-one relationships—that is, when multiple inputs map to the same output—it's possible for data to collide. Just imagine the identity problems if, say, 10 people shared the same social security number or driver's license number.

Hash functions, which map long inputs to short outputs, are important tools for efficient searching of Big Data. Unfortunately, collisions are inherent in any hash function that involves more inputs than outputs. It is possible, however, to reduce how often collisions occur. Hashes such as the message-digest algorithm (MD5) or the secure-hash algorithm (SHA) family are designed to make it infeasible to find any input that corresponds to a given output... but such hashes operate very slowly.

Fortunately, an NSA researcher has invented a system that could significantly reduce time spent on hashing calculations for large data sets. US patents 8,363,825 and 8,355,501 have been granted for a "Device for and method of collision-free hashing

for near-match inputs." The system produces more efficient hashing calculations, which in turn enable faster hashing and data retrieval. This would benefit applications such as DNA sequencing, by enabling fast searches while guaranteeing that near matches will not collide; this would make it possible, for example, to detect single nucleotide polymorphisms (i.e., genetic variation in a DNA sequence that occurs when a single nucleotide in a genome is altered).

The invention would also be useful for applications such as the construction of Bloom filters, which record whether a searcher has already viewed a particular data record. Bloom filters, in turn, can be used for data retrieval in the cloud (or elsewhere) and can be combined with homomorphic encryption for certain applications in private information retrieval.

The collision-free hashing technology is now developed, and a software demonstration is also available. To arrange a demonstration, please contact the Technology Transfer Program at tech_transfer@nsa.gov or 1-866-680-4539.



The Next Wave is published to disseminate technical advancements and research activities in telecommunications and information technologies. Mentions of company names or commercial products do not imply endorsement by the US Government.

 $This \ publication \ is \ available \ on line \ at \ http://www.nsa.gov/research/tnw/index.shtml. \ For \ more \ information, \ please \ contact \ us \ at \ at \ publication \ is \ available \ on line \ at \ http://www.nsa.gov/research/tnw/index.shtml.$ TNW@tycho.ncsc.mil.





